DOCUMENT RESUME

ED 133 733 CS 203 137

AUTHOR Nystrand, Martin
TITLE Ontological Aspects of Validity Concerns in Language Arts Assessment.
PUB DATE 76
NOTE 32p.; Report prepared at Ontario Institute for Studies in Education

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Cloze Procedure; *Communicative Competence (Languages); Composition Skills (Literary); Elementary Secondary Education; Language Arts; Reading Comprehension; *Reading Tests; *Test Construction; Testing; *Test Validity; *Writing Skills

ABSTRACT

In considering the development of language arts tests, a distinction can be made between statistical issues and ontological matters involving the objective existence and adequate characterization of the phenomenon being measured. Careful examination of standardized, norm-referenced tests and criterion-referenced tests in the areas of reading and writing indicates that, in their present forms, both types fail to meet the requirements of ontological validity. Among currently available measures in language arts, the Multiple-Choice Cloze Test of Literal Comprehension is one of the few which are related to a well-researched construct regarding comprehension. Because of the inextricable links between reading and writing, it is possible that a similar approach may be feasible in the assessment of writing skills, as well. Instructional implications deriving from the various approaches to assessment are also discussed. (AA)

**********
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
**********

ED133733

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

# Ontological Aspects of Validity Concerns in Language Arts Assessment

Martin Nystrand

The Ontario Institute for Studies in Education

1976

CS203137

Ontological Aspects of Validity Concerns in
Language Arts Assessment

Any concerted effort to specify assessment procedures and construct monitoring instruments for effective school use must certainly involve the test makers in fundamental questions regarding the nature of the area of concern. Questions which need consideration before construction of authentic measurement instruments for such use include:

(a) What is important and why?

(b) What is the character of the area or phenomenon in question?

(c) What is acceptable evidence regarding the occurence of the phenomenon?

In language arts, such seminal questions take the form of:

(a) What is good writing?

(b) What constitutes reading?

(c) How can we know when they occur?

Such questions are logically prior to any meaningful data collection and other important concerns regarding such statistical matters as validity and reliability.

In this essay, I want to distinguish between two orders of concerns relevant to the construction and development of language arts

tests (or any others, for that matter): statistical issues, meaning the capacity of the instrument regularly to collect and document the evidence; and ontological matters, involving the objective existence of the phenomenon in question, as well as adequate characterization. It is entirely possible to measure validly and reliably, and yet not know what is measured. It is possible too for the allegedly measured phenomenon to be more a result of the measuring and to have little status as an independent, objective phenomenon. Valid and reliable data without clarity regarding an objective phenomenon are results that have dubious value for classroom use and program decision-making.

An interesting example of these problems is the approach taken by the American Educational Testing Service (ETS)'s College Entrance Examination Board (CEEB) to the measurement of writing ability on the Scholastic Aptitude Test (SAT). As of 1973, ETS's assessment of writing ability was based on research indicating substantial correlations between performance on certain objective, machine-scored items and an independent, trained panel's assessment of actual writing samples. Correlations were found to be particularly substantial in items involving usage editing and sentence correction. Eight item types in all were assessed, and by manipulating them in combinations on test forms, the SAT test of Writing Ability was found to possess a validity coefficient of .7 to .8.

Such validity coefficients serve CEEB's function and mandate well. In one hour of testing, all of whose results are machine-scored, and not a difficult test to administer, literally tens of thousands of students across the U.S.--and around the world--regularly indicate to CEEB and to whomever the scores are forwarded how well they can write. Their writing abilities are accurately known.

Or are they? Their writing abilities are known in the sense that the scores are reasonably dependable: their instructors in college are more likely than not to find their writing abilities to be as CEEB has reported them. ETS's researched formats yielding reliability and validity cannot be faulted seriously for statistical competence.

There is another sense, though, in which an SAT score does not report on writing ability at all. Nowhere is there an adequate definition of writing ability provided by ETS. An SAT Writing Ability score equates directly with the ability of the test taker to correct sentences and to edit for usage, to be sure. Performance on usage and sentence correction test items correlates highly with the judgment of competent, trained readers, certainly. But what constitutes the judgment of the competent readers? What is the operating, lawful account of writing ability by which they, in their collective and reliable judgment, make their assessments? The answers to these questions are poorly known, though ETS is quick to point out that "writing ability," whatever it is,

is more than the ability to perform well on usage-editing and sentence-correction test items.

When writing ability is operationally defined by panel consensus, there is little possibility of a true definition. Indeed the fundamental question, "What is good writing?" finds the answer "Writing that the panel finds good." What does the panel find to be good writing? Good writing. According to this tautology, writers who write well get the highest score. To measure according to a tautology is in effect to measure nothing.

Tautologic tests whose validities depend on correlations possess negligible educational uses precisely because they are in violation of a most basic tenet with respect to meaningful educational measurement: Measurements bearing possibilities for affecting the phenomenon in question must report the phenomenon. The phenomenon must exist independently of the instrument that measures, and must not obtain because of the measuring.

The SAT test of writing ability, in short, is dependable for an estimation of writing ability, at least as the readers judge it. But it is dependable in the way that the appearance of salt on the table will most of the time assure us of pepper nearby, or in the way that there is a dependable relationship between a country's rum consumption and its gross national product. The validity of the SAT rests in a correlational association. If usage is good, the writing likely will be. At least this year.

That salt goes with pepper, or that rum goes with industrial production, however, does not explain or define the association. An additional shot of rum into the national arm is not what most money people would consider a major solution to economic ills. Nor does a dependable association of high scores on usage items with writing ability explain or define writing ability. At least most bright people, including the chiefs at CEEB, hope not.

Another way of examining the problem is to consider the formula involved in CEEB's computations. Writing Ability (or W.A.) for ETS = f(usage-editing performance, sentence correction, . . .). It is a formula which explains nothing more than ETS computations; it will not tell us why usage editing and sentence correction are important tokens of writing ability. Indeed, [W.A. = f( . . .)] does not truly assure us that they are at all. Good writers certainly do more than correct errors in usage, etc., well. But what that might be is unexplained by [W.A. = f( . . .)].

There is an important difference between a <u>correlational association</u> and an <u>ontological abstraction</u>. A correlational association is a sign indicating a dependable and reciprocal correspondence without accounting for causality. Salt and pepper, or [W.A. = f( . . .)], is such an association. An ontological abstraction, on the other hand, is a symbolic, lawful representation derived inductively and accounting

for causality. Laws of physics and algebraic formulae are examples of such abstractions. Correlational associations cannot serve algorithmically for purposes of prediction since only the particular combination will be found regular and dependable: sugar or flour cannot be substituted for salt in the salt-and-pepper experience. Ontological abstractions, on the other hand, will serve algorithmically. Substitutions can be made for x, y, z; p and q. Such abstractions allow for prediction and control. An ontological abstraction, in effect, purports to comment on the character of phenomena as they exist independently of observers, whereas the correlational association demonstrates only an incidental relationship.

The crispness of this dichotomy is difficult to maintain, particularly in the sense that events for humans never exist totally independently of their consciousness and perception--even for the great ontological abstractionists. It is possible to debate the real differences between the two types of relationships. For some, lawful abstractions demonstrate wholly objective relationships; for others, such abstractions are merely the multiple regressions of correlating terms allowing for substitutions. Still other investigators may choose to examine the consequences of adopting one or the other orientation, but such examinations usually result in still other ontological abstractions concerning orientations towards phenomena.

Be these debates as they may, what cannot be disputed is the existence of researchers' attempts to work out correlations and abstractions. My purpose here is to consider the consequences of both for language arts assessment.

It may legitimately be asked why CEEB need account for what makes good writers write as they do, why CEEB need account for more than a good estimation of writing ability as college instructors are likely to find it. What's wrong with a correlation? The answer, of course, is that there is nothing wrong for CEEB's purpose. What is questionable is the extent to which correlational testing serves educational purposes. If the purpose of assessment is to relate to learning in any way helpful to anyone, any test based on correlational associations is of dubious value. A diligent student could learn to perform well on the SAT, for example, without ever writing a full composition in twelve years of schooling prior to taking the SAT. Such a student might do quite well picking out the "correct errors" in ETS's test items, and yet found to be lacking in writing ability by a look at a set of elicited writing samples. If such a devious strategy existed categorically across the United States (perhaps constituting a conspiracy), ETS would undoubtedly conduct new correlational studies to assure the validity of its revised tests. Even if only usage were massively mastered, ETS would need to adjust its item-type

balance to again obtain the high validity of its tests.

There are either no instructional implications or altogether wrong instructional implications to be derived from a formula such as [W.A. = f( . . .)]. No one who writes much nor anyone who teaches writing very well could take seriously the idea that a writing program should consist in toto of the systematic and thorough perfection of usage-editing and sentence-correction skills, whatever they might be. Reputable research concludes that such an approach indeed mitigates against learning to write. [W.A. = f( . . . )], in short, misses the character of writing ability. It is a psychometric formula for ETS computations, not a principle or law in the sense of $e = mc^2$ or $s = 1/2\ gt^2$. ETS's writing ability formula in essence implies a reciprocal correspondence without accounting for a causality. An increase in the ability to edit for usage will affect W.A., but will not necessarily affect writing ability. In short, [W.A. = f( . . .)] possesses no implications for learning.

The essential requirement for any assessment that is to have implications for learning in any authentic sense is that the assessment be of an ontological nature, not a correlational one. This requirement has two corollaries. First, since the major purpose of schooling is to affect learning positively, any assessment aimed at such a purpose must be based on an <u>adequate characterization of the</u>

phenomenon in question, not on validity coefficients involving correlations rooted in tautologies. Correlations may serve oil companies well in finding oil, and they may help us wend our ways to salt at the dinner table, but when it comes to assessment for instructional purposes, schools cannot afford to be in the business of merely identifying students who bear the salient manifestations of achievement, and then reporting the winners after the fact.

Another corollary follows directly. Tests must not generate data which, if acted upon, will contribute negatively to, or make no difference in affecting the phenomenon in question. If in fixing a house, for example, we seek to level the sagging foundation of a porch, we will not accept anything less than the instrument whose measure will allow us to genuinely correct the slope. It is unlikely that we would routinely accept a qualitative analysis of the paint, or the gross weight of the front door as relevant information. Ideally, data that are useful for positively affecting phenomena are data that can be used for purposes of prediction, not just description.

Assuring that students learn to read has, of course, been one of the oldest and most venerable concerns of schooling. Assessing reading via standardized tests is a more recent but, for many, an equally venerable concern. Until recently, the universal form of standardized testing in reading has been norm-referenced. A look at

the construction of norm-referenced tests of reading is revealing. The test makers begin by generating vast numbers of test items which they feel bear on the area of investigation, in this case reading. After a large corpus of items is prepared, the items themselves are tested empirically for their power to discriminate among students. Ideally, each item is accepted or rejected on the basis of half the target population getting it right and half getting it wrong. A well-researched standardized test of reading has the power to tell how Johnny as a fourth grader compares to all other relevant fourth graders, but nowhere is there a definition illuminating such test headings as "Reading Comprehension" or "Vocabulary."

Norm-referenced tests of reading are correlational in two senses. They are correlational in the sense that an individual's performance is found to correlate in a particular way with the performance of all other students in the target population; and they are correlational in the sense that there are substantial correlations among all the major standardized tests of reading. That such a test should document individuals' performances in relation to the total group, however, does not constitute a definition of reading; and while the fact that the major standardized tests of reading correlate with each other may mean that they are essentially equivalent, the absence of any explanation for the correlation leaves as undetermined what they all measure equivalently. It is curious that

standardized tests of reading also correlate highly with tests of verbal IQ, but such a finding simply raises more questions than it answers.

In any event, with such a lack of clarity concerning what is being measured, standardized tests of reading would seem to have an almost nonexistent ontological status with regard to the phenomena they purport to measure. Their validities are to be found in their power to discriminate among people, not in their capacities to comment on an individual's reading according to an adequate characterization of the phenomenon in question. Because their status is correlational rather than ontological, standardized tests of reading can play no role in the school's purpose of assuring that children will learn to read. If anything, the opposite is true: the requirements of a normal distribution which are a part of any standardized, norm-referenced test guarantee the users that 50% of all those who take the test will be found to be "inadequate readers."

A somewhat current attempt to deal with many of the inadequacies of norm-referencing is an effort called criterion-referencing. The criterion-referenced test developer is not generally interested in comparing one student to a large population. The criterion-referencer is more interested instead in specifying the achievement of individuals. Although criterion-referencing is still too new to

have taken a universal form (at least in the way that norm-referencing has), the general approach is to begin with the delineation of comprehensive lists of objectives important to the users (adults in authority), and then to state these objectives in terms of specified behaviors. The latter process is usually referred to as operationalization. Several test items are then generated for each objective. The intent of the criterion-referencer, in short, is to comment specifically on individual achievement with reference to objectives, not other students. The essential query is the extent to which individual students have achieved the objectives which have been laid out by the school.

A criterion-referenced test can be correlational, but need not be. One method of criterion-referenced test development is to set as individual objectives each of the performances used in the standardized test to identify and rank students of varying abilities. If ETS, for example, has found usage editing and sentence correction to be high import behaviors of good writers, usage editing and sentence correction can each be set as individual objectives to be mastered and measured. It is entirely possible to categorize various subtypes of sentence correction in detail (e.g., commas in series, semi-colons, capitalization of proper names, etc. ad infinitum), and then proceed to write objectives and test items for each. Unlike

norm-referencing, criterion-referenced test items are not accepted or rejected on the basis of their power to discriminate and assure a normal distribution; their use is determined mainly by the extent to which they accurately measure stated objectives. The criterion-referencer, in short, would be likely to include test items which 98% of all students might get right as long as the test items measured an important objective or set of objectives; the smart norm-referencer, on the other hand, would reject such items.

Such a criterion-referenced test in language arts is clear about what it measures in the sense that the objectives are available to anyone for inspection. Criterion-referenced tests, furthermore, need not be correlational in the sense that the behaviors they measure are the salient traits of high achievers only. Yet such a shift in approach still does not assure the ontological character of the test. If norm-referencing has ontological difficulties related to the use of correlational associations rather than lawful abstractions, a major hazard in the development of criterion-referenced tests in language arts is the ease with which one particular assumption can be made: the assumption that a substitution of specified objectives for correlational associations as the modus operandi for test development will necessarily provide the users with authentic achievement data, as well as results bearing genuine implications

for assisting the learner to learn better.

An ad hoc collection of socially validated and specific objectives simply does not constitute an adequately researched theoretical statement regarding a known phenomenon. There is a literal infinitude of things that can be identified and counted in language, and the degree to which they relate to the phenomenon in question is by no means assured by the matching criteria of consensus as to their importance within a writing group along with adequate specificity.

To weigh any criterion-referenced test in language arts for ontological status, it is necessary to consider the second corollary regarding the usefulness of the data. Essentially the question concerns the effects on learning resulting from acting on the data provided. Several currently available criterion-referenced tests of reading report detailed profiles on the adequacy of students' mastery of such items as consonant blends, dipthongs, sight words, structural markers and various aspects of syntax, as well as comprehension objectives concerning main ideas, inferences, and the like. If all of these headings in fact constituted the various, relevant, and sundry components of reading ability; and if fluent reading were achieved by adequate instruction in each and all, the logic and desirability of detailed profiles on all of these headings might be unquestioned.

Yet a recent study in New York State (O'Reilly, 1975) concludes that a categorical increase by a factor of four in all types of reading instruction available throughout the state over the period of a year made no significant difference by anyone's measure when increases in comprehension were examined. On the other hand, significant increases were uncovered when resources were allocated to classroom libraries containing books that students could and would read.

This research has merit on its own, particularly insofar as it underscores the importance of the notion learning to read by reading (Smith, 1971). Compared with other research on language and learning from other quarters, though, it has particular bearing on the present discussion. In Research in Written Composition (1963), Braddock and his associates concluded in a review of major studies of the effects of instruction in formal grammar on achievement in writing: "the teaching of formal grammar has a negligible or, because it usually displaces some instruction and practice in actual composition, even a harmful effect on the improvement of writing." (pp. 37-38). In independent research, psycholinguists Katz and Fodor define comprehension as the ascertainment of "grammatical and semantic relations which obtain within and among sentences of the discourse" (1963, p. 172). These three studies, from the areas of reading, writing, and psycholinguistics, are currently among a great number which

continue to support a conception of language as social behavior, an event involving the construction of relationships and combinations by individuals for the purpose of reducing uncertainty about themselves and the world. The key words in the conception are relationships and use. Words seem to have very little meaning without a consideration of how and with what other words they are used, and the meaning of any combination is not equivalent to the sum of the componenents. The business of learning language is particularly impervious to instruction which treats language as a discrete body of knowledge for the purposes of explicit mastery. In an important sense, the more language is divided, the less anyone seems to conquer anything.

Given current schools of thought on language, the above represents an all too brief summary of the ontology of language as an objective phenomenon. As with any research, confidence in the formulation of the phenomenon is increased as empirical evidence accumulates from independent studies conducted by researchers who do not collaborate. A series of confirmed hypotheses is usually prelude to theory, a general framework suggesting lawful relationships and purporting ontological claims.

So far I have postulated ontological abstractions as the prime requirement for any test bearing implications for learning. This

postulate was followed by two corollaries:

1. adequate characterization of the phenomenon in question as demonstrated by confirmed hypotheses from independent studies and suggesting lawful (i.e., algorithmic) relationships.
2. specificed procedures for the generation of data which can be used for the purposes of prediction and control.

Standardized, norm-referenced tests regarding reading and writing must be dismissed as such tests because their validities rest in correlations rather than abstractions. Criterion-referenced tests involving eclectic, ad hoc collections of objectives and test items fail to meet the requirements because of the essential incompatibility of excessive fragmentation with the nature of language, as well as the lack of psychological interrelatedness among the objectives.

Are there any currently available measures in language arts which qualify? In this period of rapid development in the area of measurement and evaluation there is one in particular that deserves consideration and close study. It is the Multiple-Choice Cloze (MCC) Test of Literal Comprehension, developed by Robert O'Reilly* and his

* Robert O'Reilly is currently Director of Research and Evaluation with the Montgomery County Public Schools in Rockville, Maryland..

associates at the Bureau of School and Cultural Affairs, the New York State Department of Education in Albany.

The development of this test is reported in a number of recent papers and monographs, most notably in a 1975 monograph entitled SPPED Cloze Exercises in a Multiple-Choice Format. Citing a great number of theoretical and empirical studies regarding comprehension and reading process, including those of F. Smith, J. Bormuth, and Katz and Fodor, the New York group has converted the original use of cloze as a test of readability into a test of reading comprehension. The following is a typical passage with accompanying test items:

THE YOUNG WHALE

The young whale tapped his teeth and _____ Coos Bay. He had been _____ in January, a magnificent _____ of sixteen feet. Upon his _____ in the whale world, he had been _____ nuzzled by his giant _____, who, without arms or _____ with which to hug him, _____ her love by circling him. She _____ him to the surface to _____, then, tapping her body, she showed him where he would find her milk.

a. circled
b. loaned
c. obeyed
d. became
e. farmed

a. thankful
b. nervous
c. slow
d. foul
e. born

a. hawk
b. quail
c. pipe
d. male
e. flea

a. scorn
b. location
c. raccoon
d. blister
e. arrival

a. fern
b. lap
c. puppet
d. beech
e. mother

a. sauces
b. feet
c. cuts
d. hills
e. inns

a. computed
b. decorated
c. copied
d. expressed
e. repaired

a. stitched
b. married
c. glued
d. led
e. lit

a. indignantly
b. immediately
c. warily
d. hoarsely
e. viciously

a. ache
b. bow
c. blow
d. add
e. fade

The multiple choices listed under the passage include the original, deleted words, along with distractors which compete syntactically but not semantically (there are no synonyms). In all cases, only nouns, verbs, adjectives, and adverbs are deleted, and the deletion rate is increased as readers progress. The essential task of the test taker is to reconstruct the original text in its full coherence by working back and forth between the broken text and the multiple choices.

Much of the significance of the MCC is to be found in the solutions it offers to difficult distinctions which have been attempted in the past. One such distinction is that between "explicit" and "implicit" comprehension. A major assumption involved in cloze testing is that for the beginner, all is implicit, hidden. A major reading objective is for the reader to render increasingly more about print increasingly explicit, an objective which the MCC measures directly.

The MCC offers a significant measurement solution, too, to problems involved in some criterion-referenced efforts to provide achievement data based on the leveled components of neat but psychologically unfounded reading taxonomies. There is no attempt in the MCC, for example, to provide profiles on "sequenced achievement in I. sound-symbol relationships; II. whole words and vocabulary; III. sentences and syntax; and IV. passages." The MCC is based instead on a construct which stresses the interrelationships and mutual dependencies of words between and among each other. The elements of language are conceived as necessary but insufficient to account for comprehension.

The Test Development Notebook which accompanies the 1975 monograph contains detailed test specifications in algorithmic form, meaning essentially that the New York group has not only standardized its test items, but significantly, has standardized procedures for generating test items as well. New passages and items are easily added.

The MCC is particularly significant insofar as its procedures relate to a well-researched construct regarding comprehension. This construct is that of Katz and Fodor: comprehension is the ascertainment of semantic and syntactic relationships between and among words. Involved is the gestalt notion that the whole is not equivalent to the sum of the parts. As Smith notes, meaning is in the reader, and

the query of the MCC, with its systematically broken texts, is not entirely unlike another inquiry: How far must a viewer draw back from a blown up newspaper photo--only dots--before a meaningful representation is found?

How much confidence can the users of the MCC have in the test's power to measure comprehension? Although this question continues to be a major source of research for the developers of the test, the initial validation studies (O'Reilly, Schuder, Kidder, 1976) are positive. After examination of a great number of tests of comprehension, norm-referenced and others, hypotheses were developed regarding expected correlations between the MCC and the other tests. It was predicted, for example, that the MCC would correlate highly with Bormuth's Wh- Item Test of Literal Comprehension since both tests access the same syntactic and semantic features: the focus of the MCC's deletions on nouns, verbs, adjectives, and adverbs is essentially the same as Bormuth's Who, What, Where, When, How. This prediction was also made on the basis of the relative precision of the two tests to measure a single trait specifically related to reading and distinct from verbal IQ, a topic explored by Carroll (1972). On the other hand, only moderate correlations were hypothesized between the MCC and standardized, norm-referenced tests after a study indicated that the latter measured a great number of

mixed, poorly defined traits. Empirical confirmation of these hypothesized correlations was taken as initial support for both the Katz and Fodor construct and the MCC as a measure of comprehension so defined.

Support for the construct validity of the MCC would in many respects seem to be broader, however. There is a substantial suggestion of confirmation inherent in the coalescence of the Braddock studies of writing, and the New York studies of the effects of instruction on achievement in reading. These pieces of research serve essentially to heighten the importance of relationships and use as appropriate aspects of language learning. Considered alongside the considerable illuminations of gestalt psychology regarding meaning--the parts necessary but not sufficient to account for meaningfulness; meaning in the individual--the MCC would seem to underscore the importance of the individual reader's role in attributing meaning or significance rather than detecting it fully formed. Meaning and comprehension involve active construction, not passive reception. They are part and parcel of the entire experience of learning to read, not advanced aspects to be dealt with after mastery of "fundamentals." The ontology of the MCC is substantial and represents a rejection of the inadequate approaches of idealism and realism to meaning, positing more a structuralist solution instead.

Because of an inextricable link between reading and writing, it seems entirely possible that the work of the New York group in the measurement of comprehension will ultimately have spinoffs in the assessment of writing, as well. This possibility is currently being explored in the Language Arts Assessment Project of the Ontario Institute for Studies in Education. A major question in the assessment of writing has always been: What is good writing? As part of the formulation of the CEEB test of Writing Ability, Paul Diederich conducted an investigation to answer this question. In this study, he made multiple copies of student writing samples, and then distributed each paper to a great number of readers, not all of whom were teachers, for reading and marking. Virtually every paper so read received every possible mark, from Superior to Failing. A factor analysis of comments written on the papers revealed five basic clusters among the judgments: ideas, mechanics, organization, style, and spelling. When panels of readers were then taught to be explicit about what they valued, and trained to be consistent in applying the agreed-upon criteria, highly reliable readings of papers giving a normal distribution were obtained.

Such a procedure substantially improved the reliability of the judgments, and the articulation of the criteria for marking increased the validity, but no study was apparently ever conducted to account

for the relationships between the agreed-upon criteria and writing ability in any causal sense. The Diederich technique, in short, sorted out a statistical nightmare but did little to illuminate writing ability according to lawful abstractions.

Considering what is known concerning the nature of language, there is reason to suspect that Diederich's original finding regarding highly vacillating judgments of readers was closer to an actual account of writing ability than the ultimate normal distribution of the trained panels. From the time a child learns to form the letters of the alphabet, there is an important aspect of audience involved in any writing. Competence in writing specifically requires awareness of the needs of the reader on the part of the writer. As a minimum the reader must be able to make sense of what gets written. This stipulation would seem to come closer to a definition of writing ability than a set of isolated and fixed criteria that the writer is asked to match. Writing is very much social behavior involving writers with their readers, and many of the criteria of good writing reside in the readers, not the text. Writers regularly stand or fall to the extent that they control shared, relevant terms of expression.

Given this definition of writing ability, it may be assumed that equivalent readers will make equivalent sense of a given text. Cloze is one procedure for making such an assessment; its focus is on the space between writer and reader. The writer either specifies his

audience, or an audience is specified, and the cloze score of the audience within relevant time constraints may be taken to indicate the success of the writer to make sense for his readers. Current research on cloze may well have implications for more than reading.

Because the MCC is ontological, it has substantial implications for teaching and learning. Its data can be used for prediction and control since we can have confidence in its capacity to document an independent, objective phenomenon. The MCC affirms the notion of meaning as the appropriate emphasis from the very beginnings of reading. To cite the construct, the teacher's role is to assist the learner ascertain increasingly more semantic and syntactic relationships. In other terms, the teacher's task is to facilitate the reader's attempts to render the implicit explicit. The teacher must understand and remember, of course, that what may be explicit to the teacher as a fluent reader may not be so to the beginner. There are classroom implications inherent in the MCC which confirm the seminal importance of the reader's ascertaining relationships by dealing with relationships, not fragments. Language is learned not by the teacher parcelling out its elements systematically, even on an individual basis, but rather by assisting the reader to come to grips with its wholeness.

There are no linear, diagnostic implications from a low cloze

score. There are, however, some awesome reminders in the test and its conception concerning what language learning requires. The essential task in using the MCC instructionally involves less questions of direct intervention to deal with the inadequate presence of various said-components of "the system" (e.g., remediating phonics), and more a challenge to those in charge of teaching reading to come to grips with what meaning is, why it must be an essential focus at all levels of learning, and why children learn to read by reading. The instructional implications of cloze are nothing short of a call for an adequate understanding of language and its learning on the part of those in charge.

The data of the MCC allow for prediction and control in the sense that they stipulate the essential requirements for learning to read. The data are wholly consistent, for example, with the prediction that the less students read or the more they are taught about reading at the expense of time spent reading, the less likely they will learn to read.

Compared to currently available norm-referenced and criterion-referenced instruments in language arts, the MCC is something of a departure. Unlike the former, it is not based on empirically derived correlations without a definition of terms. It compares to criterion-referenced tests in the sense that it defines its focus,

but allows users to have more confidence in the sense that it defines the parameters and features of an objective phenomenon within a well-founded framework. Because of its ontological, rather correlational status, it bears authentic possibilities for contributing positively to learning if it is used with understanding.

Bibliography

Braddock, Richard, et al. Research in Written Composition. Urbana Illinois: National Council of Teachers of English, 1963.

Carroll, J. "Defining Language Comprehension: Some Speculations." In Language Comprehension and the Acquisition of Language, edited by R. Freedle and J. Carroll. New York: John Wiley, 1972.

Diederich, Paul B. Measuring Growth in English. Urbana, Illinois: National Council of Teachers of English, 1974.

Godshalk, Fred; Swineford, Frances; and Coffman, William E. The Measurement of Writing Ability. Research Monograph, number six. New York: College Entrance Examination Board, 1966.

Katz, J., and Fodor, J. "The Structure of a Semantic Theory." Language 39 (1963):170-210.

New York State Department of Education, Bureau of School and Cultural Affairs. SPPED Cloze Exercises in a Multiple-Choice Format. Albany, New York: State Department of Education, 1975.

Nystrand, Martin. "Briefing Paper on 'Implicit Comprehension'" (unpublished)

Nystrand, Martin. "Notes towards an Account of Written Communication: Apologia Tendentia." (unpublished)

Bibliography / Continued

O'Reilly, Robert P. "The Contributions of Quantity and Quality of Instruction to Reading Programs." Paper presented at Productivity in Reading Programs symposium, The American Educational Research Association, 1975, in Washington, D.C.

O'Reilly, Robert P.; Schuder, R.T.; and Kidder, Steven J. "Validation of a Multiple-Choice Cloze Test of Literal Comprehension: Summary Report." Paper presented at the annual meeting of the National Council on Measurement in Education held in conjunction with the Annual American Educational Research Association, San Francisco, 1976.

Smith, Frank. Understanding Reading. New York: Holt, Rinehart, and Winston, 1971.